An Overview of Atmospheric Data Assimilation

3rd International Symposium on Integrating CFD and Experiments in Aeronautics

Capt. Paul Homan
Maj. Bob Wacker
USAFA/DFP
20 Jun 07

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUN 2007**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **An Overview of Atmospheric Data Assimilation**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Air Force Academy, USA**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
Third International Symposium on Integrating CFD and Experiments in Aerodynamics, June 2007, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**37**	

Overview

- ***Purpose: Convey how weather forecasting models incorporate diverse observational data***

- **Governing Equations**
- **Introduction to Meteorological Observations**
- **Introduction to Numerical Weather Prediction (NWP)**
- **Methods of Data Assimilation**
  - **Optimal Interpolation (OI)**
  - **3-Dimensional Variational Assimilation (3DVAR)**
  - **4-Dimensional Variational Assimilation (4DVAR)**
  - **Kalman Filter**

Atmospheric Governing Equations

- **Conservation of Momentum**

$$\frac{d\mathbf{U}}{dt} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \mathbf{F_r} - 2\mathbf{\Omega}\times\mathbf{U}$$

Total acceleration of air parcel relative to Earth surface (non-inertial frame due to Earth rotation)

Pressure gradient acceleration

Acceleration due to gravity; includes effect of centrifugal force due to Earth rotation

Viscous acceleration; neglected above planetary boundary layer; parameterized within PBL

Coriolis acceleration (horizontal component 90° right of velocity in NH)

Atmospheric Governing Equations

- **Conservation of Momentum**

$$\frac{d\mathbf{U}}{dt} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \mathbf{F_r} - 2\mathbf{\Omega}\times\mathbf{U}$$

Geostrophic Balance: Coriolis acceleration balances pressure-gradient acceleration in large-scale flow; mass field determines velocity field

$$\frac{du}{dt} - \frac{uv\tan(\phi)}{a} + \frac{uw}{a} = \boxed{-\frac{1}{\rho}\frac{\partial p}{\partial x}} + F_{rx} + \boxed{2\Omega v\sin(\phi)} - 2\Omega w\cos(\phi)$$

$$\frac{dv}{dt} + \frac{u^2\tan(\phi)}{a} + \frac{vw}{a} = \boxed{-\frac{1}{\rho}\frac{\partial p}{\partial y}} + F_{ry} - \boxed{2\Omega u\sin(\phi)}$$

$$\frac{dw}{dt} - \frac{(u^2+v^2)}{a} = \boxed{-\frac{1}{\rho}\frac{\partial p}{\partial z}} - \boxed{g} + F_{rz} + 2\Omega u\cos(\phi)$$

Hydrostatic Balance: gravity balances vertical pressure-gradient in large-scale flow; temperature determines mass field

Atmospheric Governing Equations

- **Conservation of Mass: Continuity Equation**

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{V})$$

- **Conservation of Energy: Thermodynamic Equation**

$$\underbrace{\dot{Q}}_{\text{Rate of diabatic heating per mass}} = \frac{d}{dt}\underbrace{\left[\underbrace{\frac{(u^2+v^2+w^2)}{2}}_{\text{KE per mass}} + \underbrace{\Phi}_{\text{PE per mass}} + \underbrace{c_v T}_{\text{Internal energy per mass due to temperature}} + \underbrace{p\alpha}_{\text{Internal energy per mass due to pressure}}\right]}_{\text{Total energy of air parcel}} - \underbrace{\alpha\frac{dp}{dt}}_{\text{Adiabatic heating due to vertical motion}} - \underbrace{\mathbf{V}\cdot\mathbf{F}_{\mathbf{r}}}_{\text{Work done by friction}}$$

Atmospheric Governing Equations

- **Equation of State**

$$p = \rho RT$$

- **Conservation of water vapor mixing ratio**

$$\frac{\partial(\rho q)}{\partial t} = -\nabla \cdot \left(\rho \mathbf{V} q\right) + \rho(E - C)$$

Total amount of water vapor in a parcel is conserved as the parcel moves except where there are sources (evaporation E) and sinks (condensation C)

Equations Summary

- **Seven Equations with seven unknowns**
  - ***u, v, w, T, p, ρ, q***

$$\frac{du}{dt} - \frac{uv\tan(\phi)}{a} + \frac{uw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + F_{rx} + 2\Omega v\sin(\phi) - 2\Omega w\cos(\phi)$$

$$\frac{dv}{dt} + \frac{u^2\tan(\phi)}{a} + \frac{vw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} + F_{ry} - 2\Omega u\sin(\phi)$$

$$\frac{dw}{dt} - \frac{(u^2+v^2)}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + F_{rz} + 2\Omega u\cos(\phi)$$

$$\frac{\partial \rho}{\partial t} = -\nabla\cdot(\rho\mathbf{V})$$

$$\dot{Q} = \frac{d}{dt}\left[\frac{(u^2+v^2+w^2)}{2} + \Phi + c_v T + p\alpha\right] - \alpha\frac{dp}{dt} - \mathbf{V}\cdot\mathbf{F_r}$$

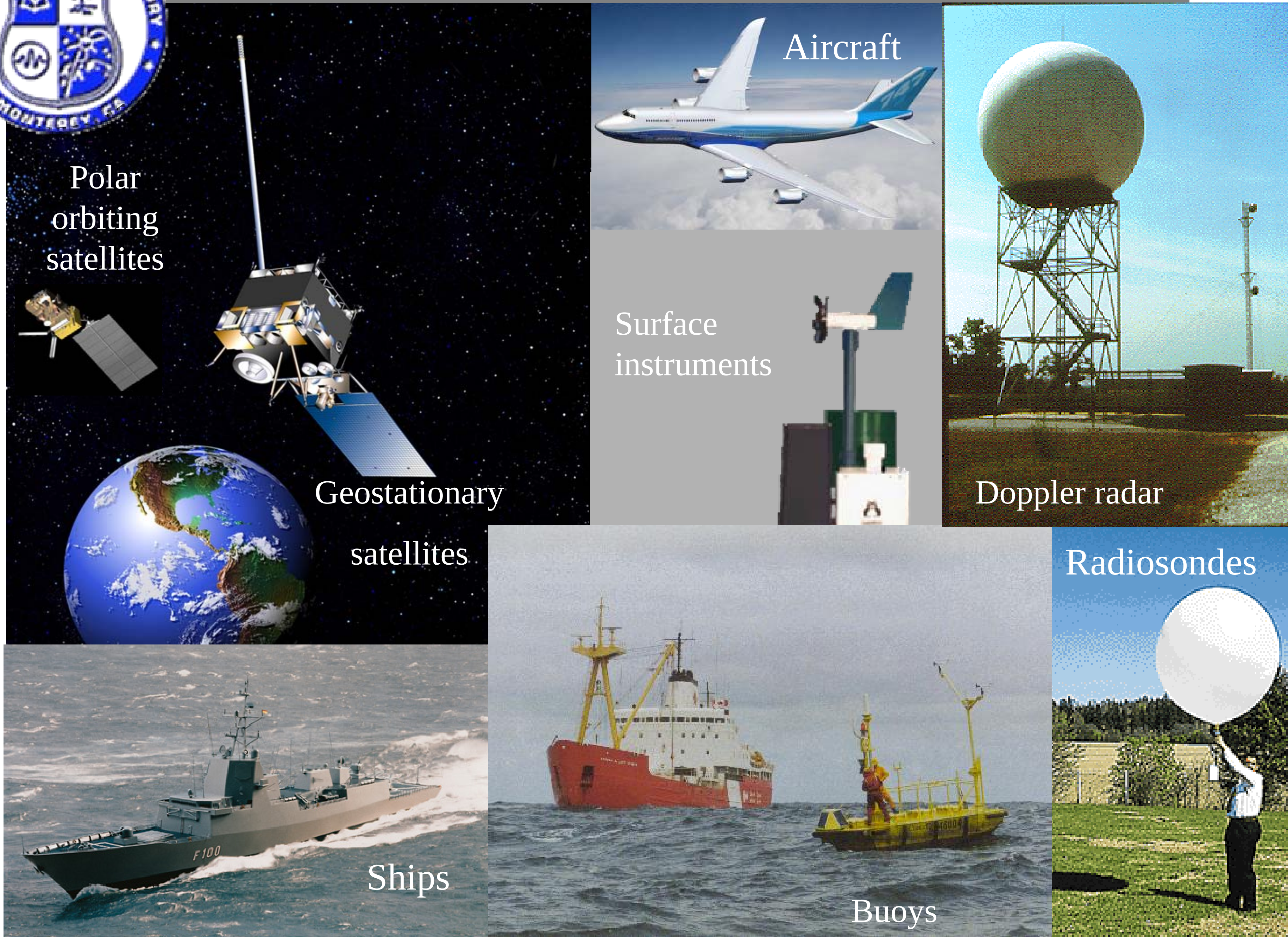
$$p = \rho RT$$

$$\frac{\partial \rho q}{\partial t} = -\nabla\cdot(\rho\mathbf{V}q) + \rho(E-C)$$

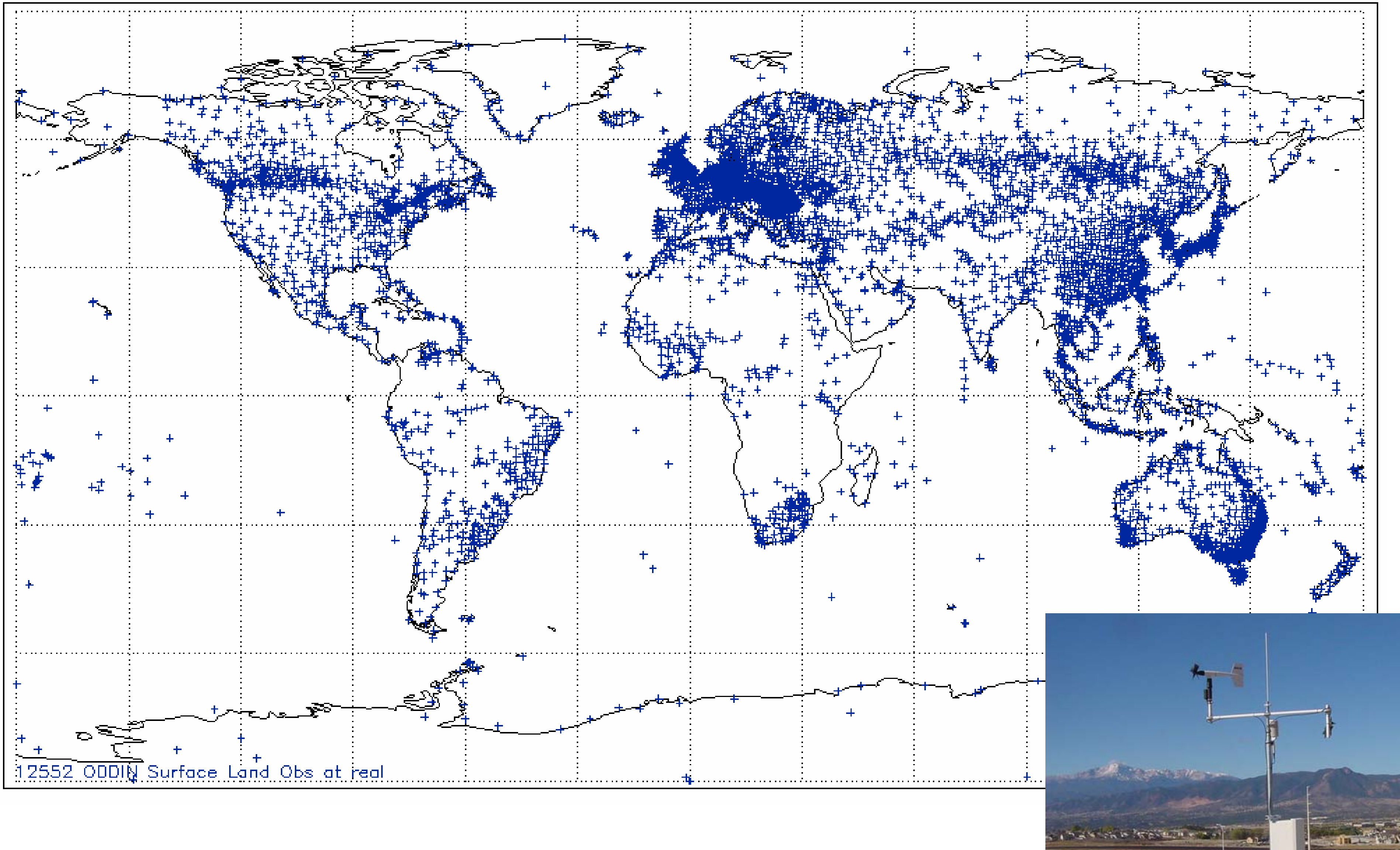
Introduction to Meteorological Observations

- In situ
  - Surface
  - Upper-air
  - Buoys
  - Aircraft
  - Ships
- Remotely sensed
  - Radar
  - Wind Profiler
  - Satellite soundings
  - Satellite winds
  - May not be meteorological variables (i.e. radiance, reflectivity, etc)

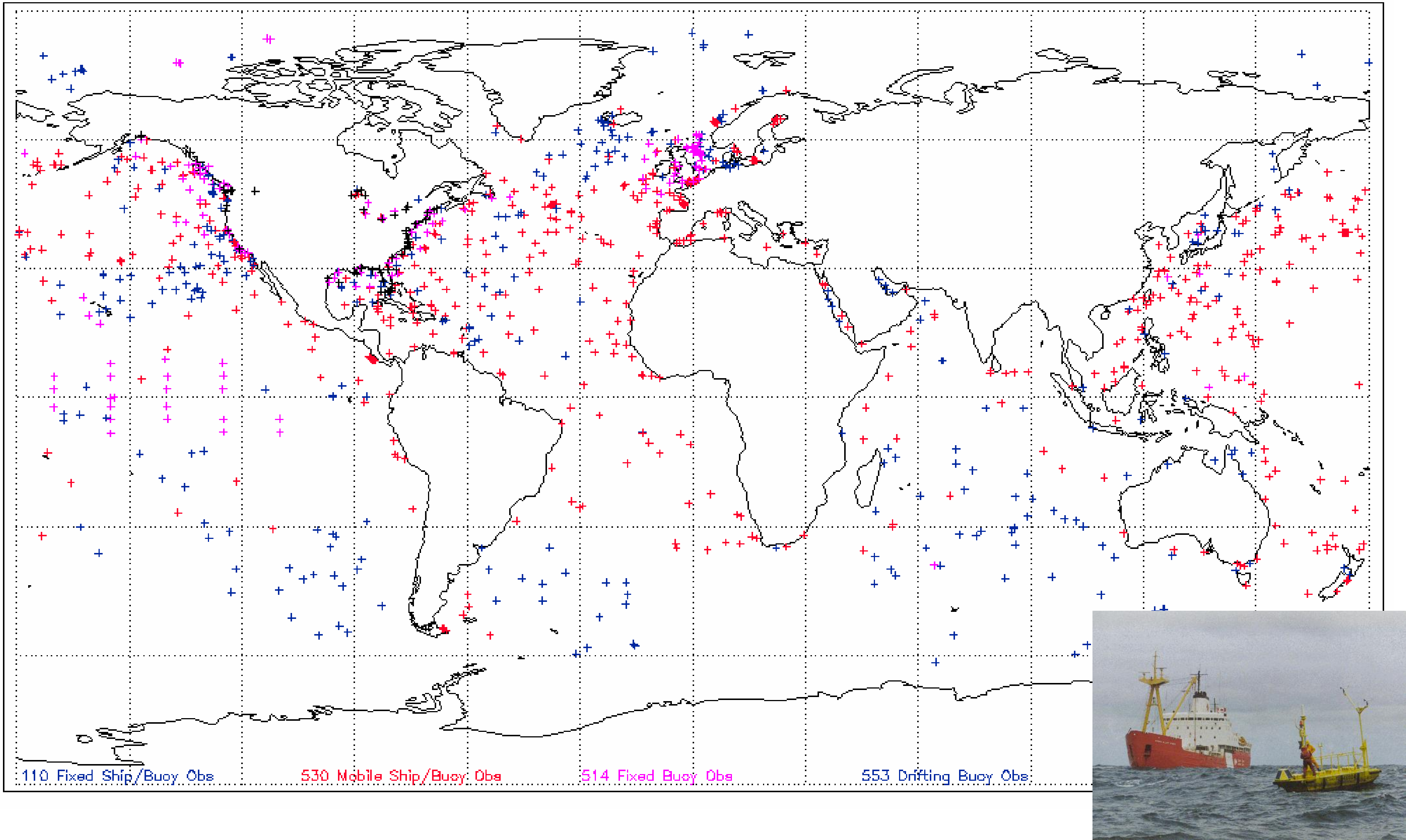
Weather Data Observing Platforms



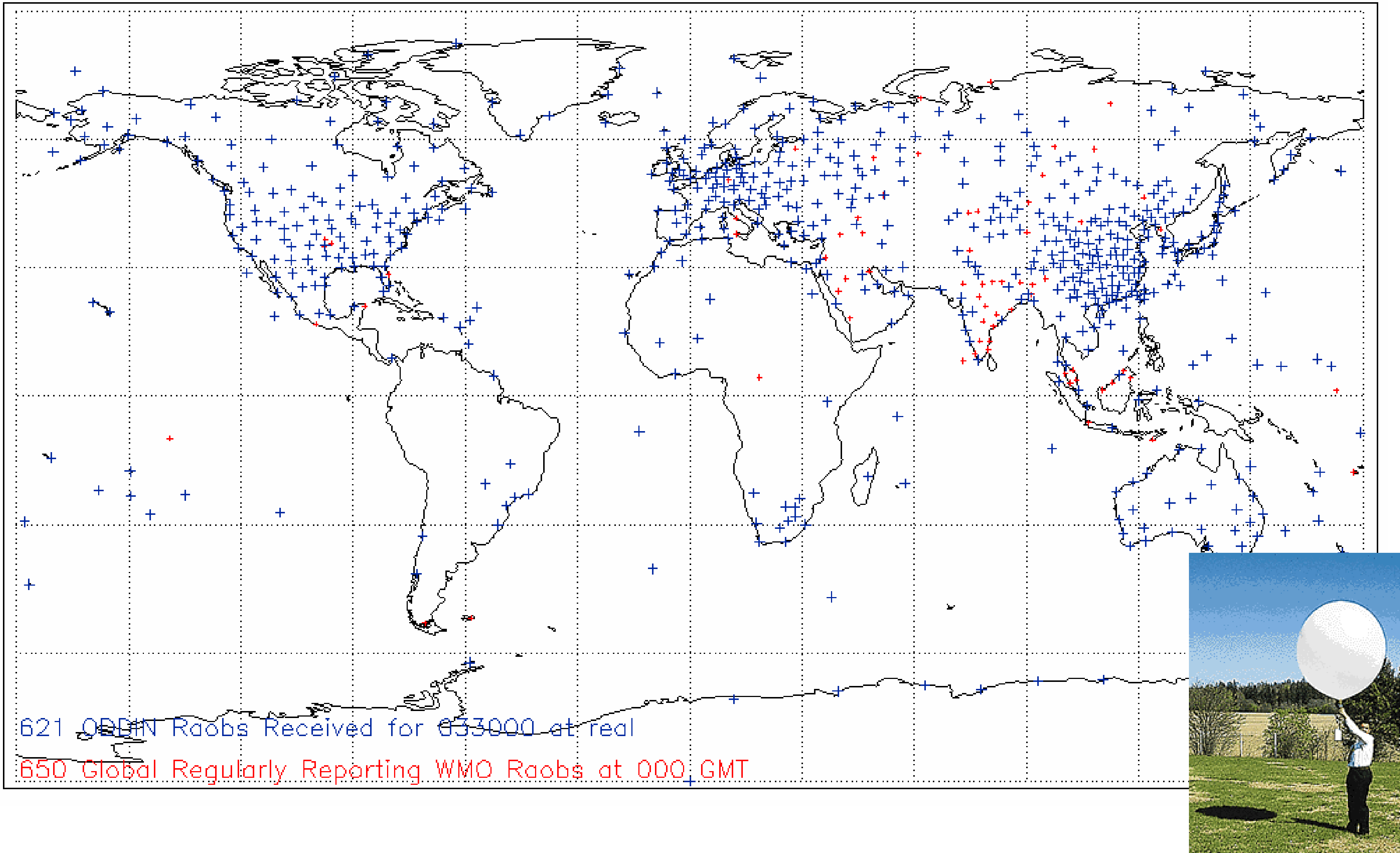
Global Surface Land Observation Coverage



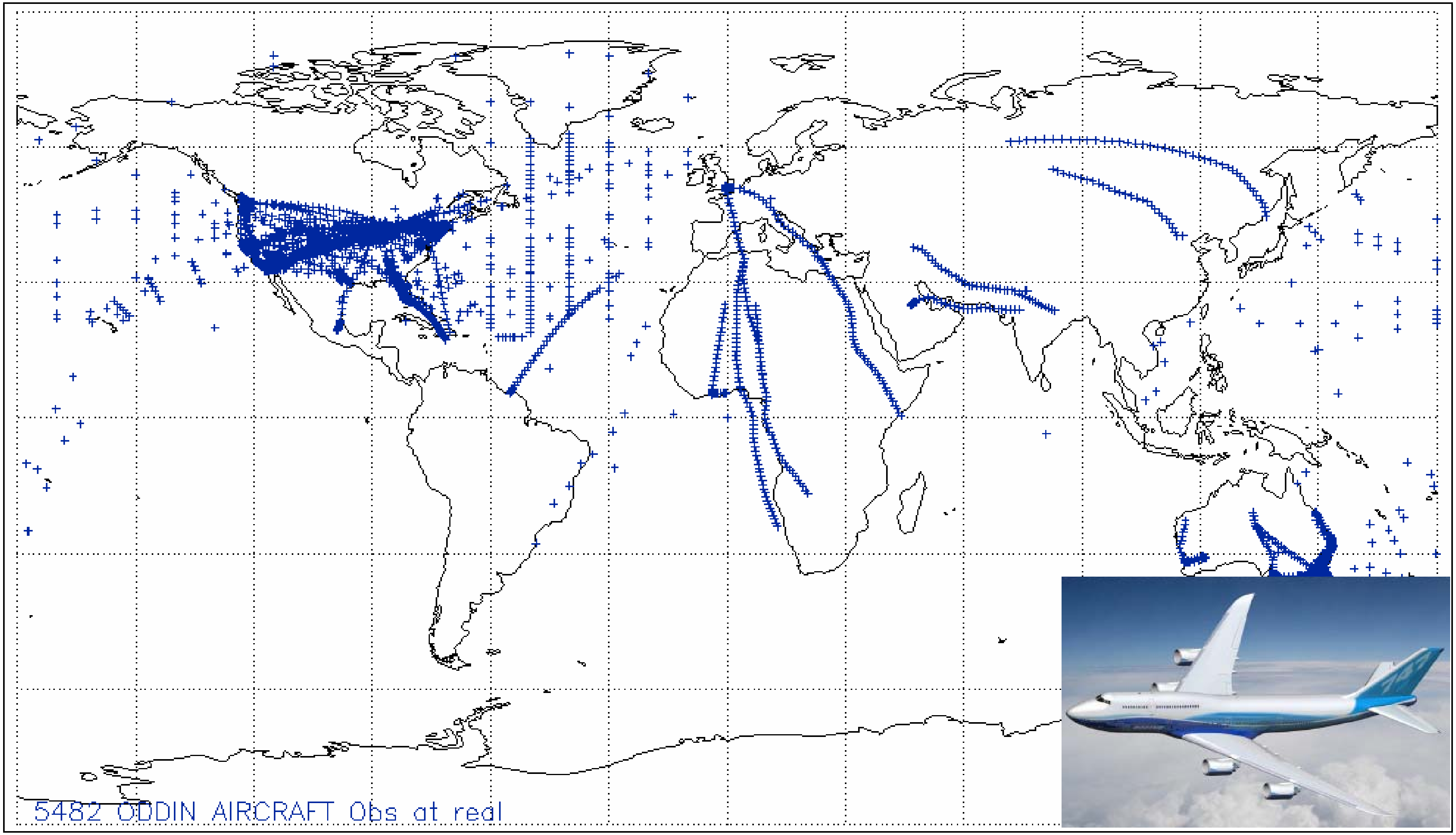
Global Surface Ocean Observation Coverage



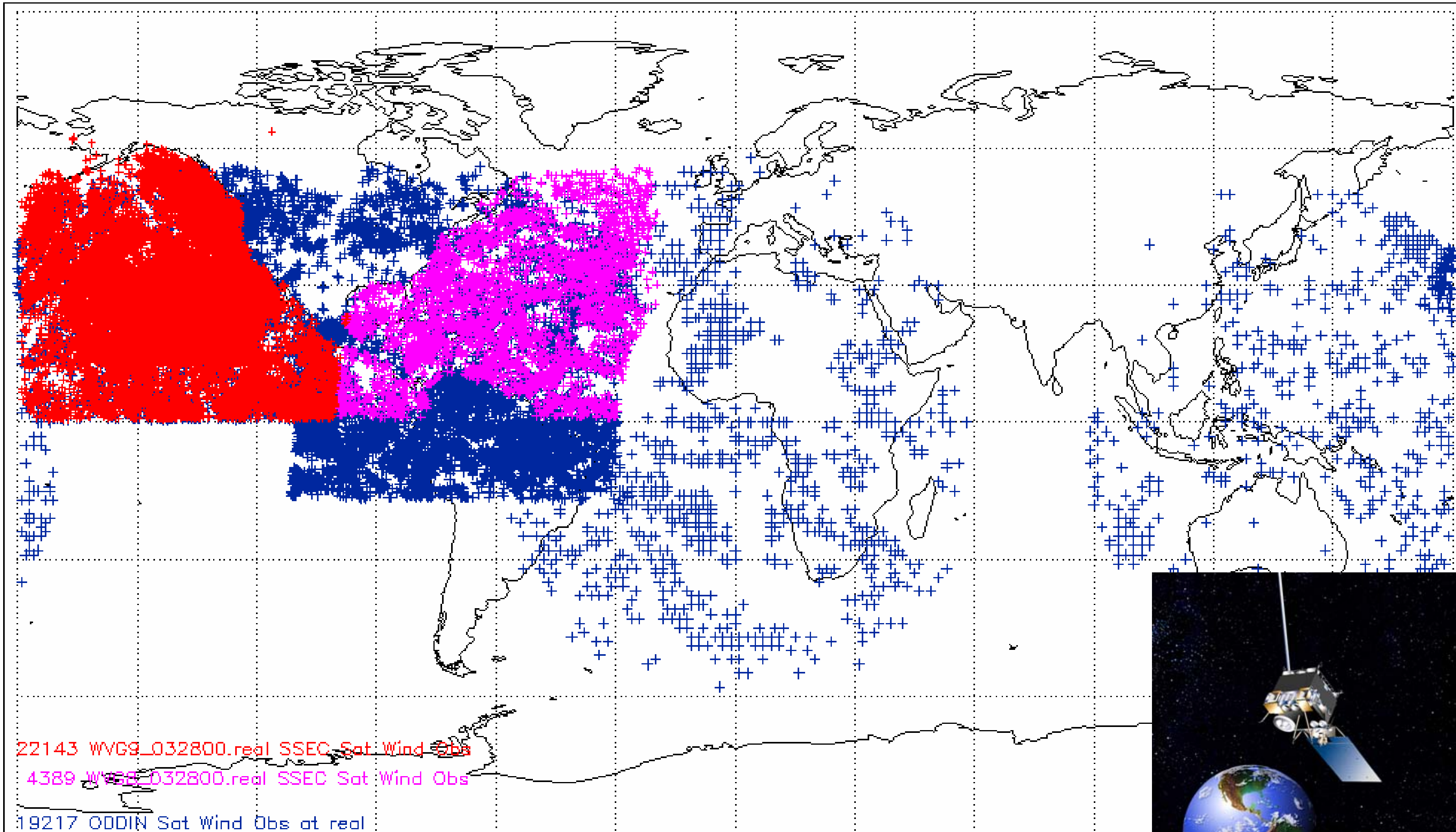
Global Radiosonde Coverage



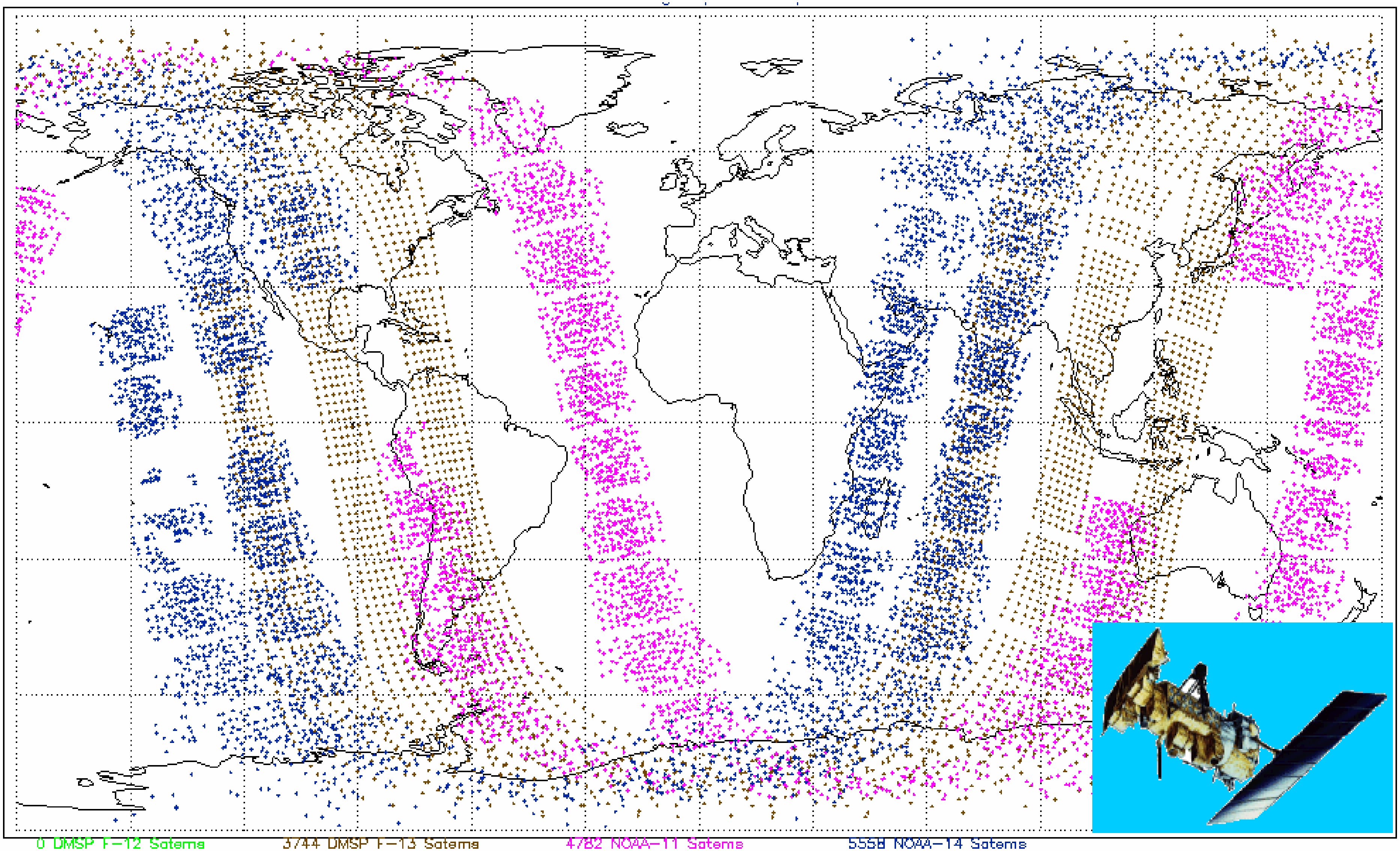
Global Aircraft Coverage –Typical 6 hour period



Global Satellite Wind Coverage



Polar Orbiting Satellite Temperature Sounding Coverage



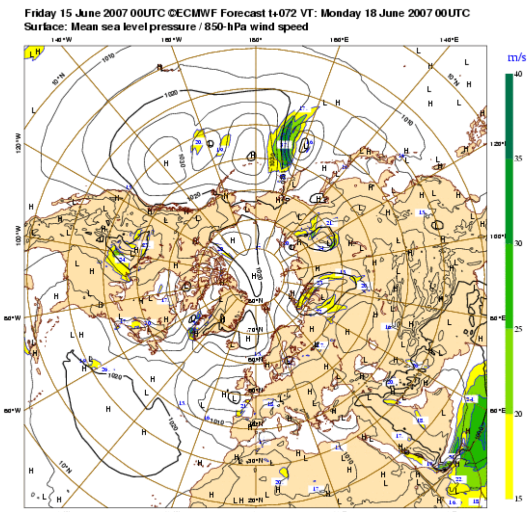
Introduction to Numerical Weather Prediction

- **Global models- Spectral**
  - **Spherical harmonic representation of data**
  - **Equations of motion in amplitude space**

  - **ECMWF- European Center for Medium-Range Weather Forecasts**
    - **T799L91- Horizontal grid length of 25km and 91 vertical levels**

  - **GFS- Run by NCEP (NOAA)**
    - **T382L64- Equivalent to about 40km resolution with 64 vertical levels**

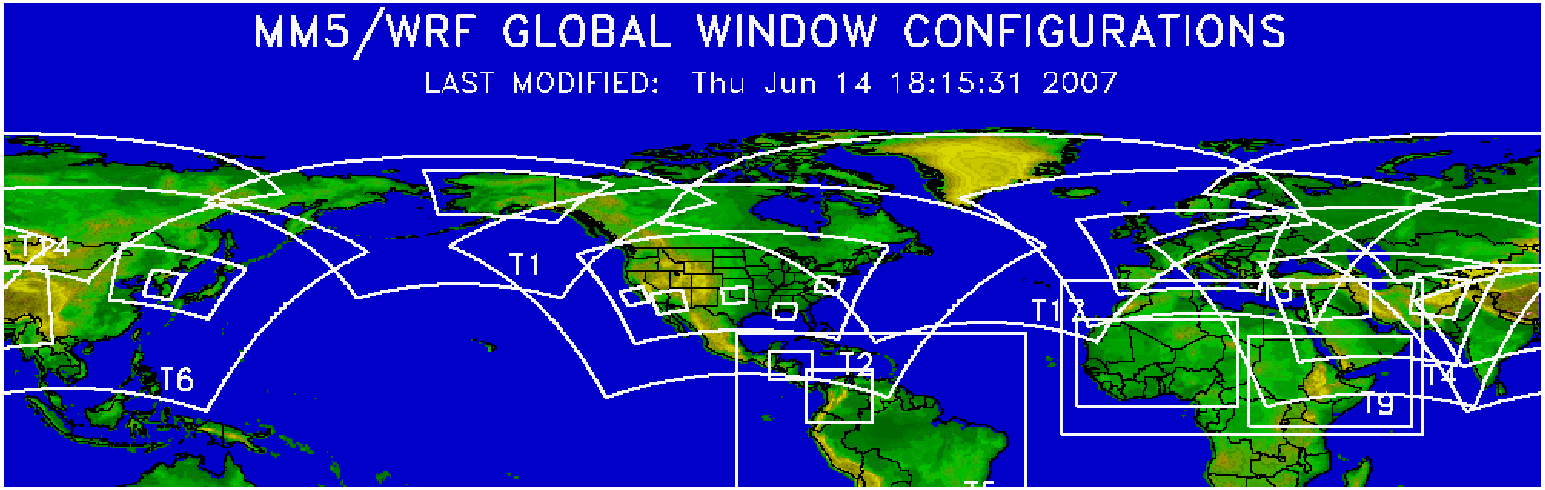
  - **NOGAPS- Fleet Numerical (Navy's Global Model)**
    - **T239L30- 55km with 30 vertical levels**



http://www.ecmwf.int/products/forecasts/d/charts/medium/deterministic/

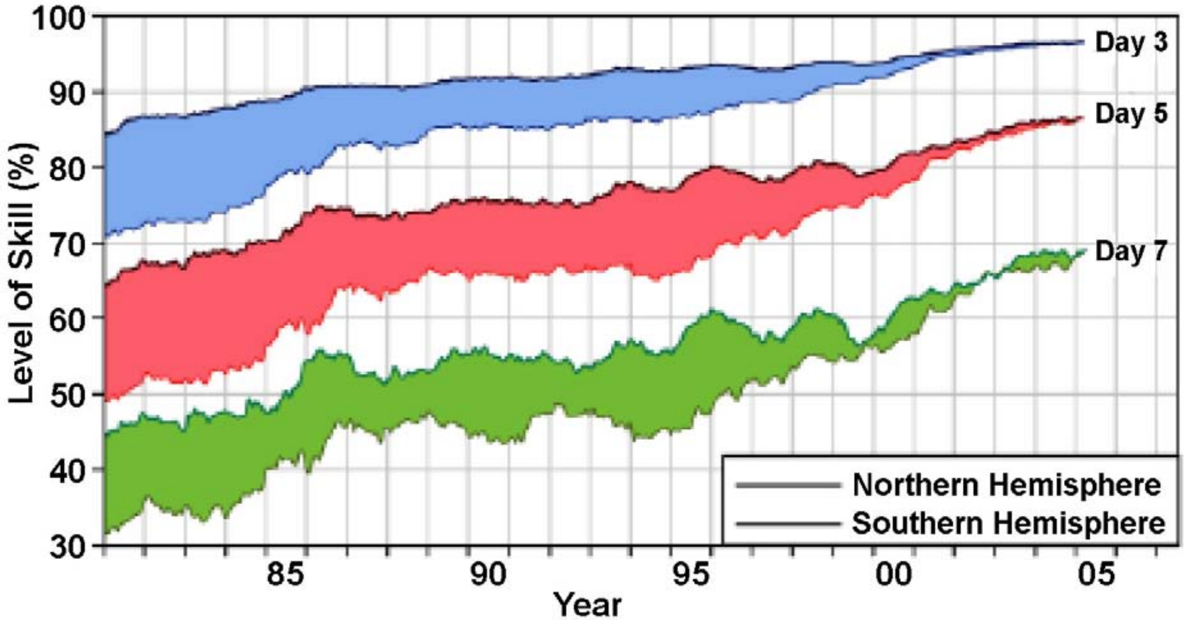
Introduction to Numerical Weather Prediction

- Limited Area (Regional) Models- Finite Difference
  - Nested within larger global or regional models
  - Boundary conditions via global model or bigger regional model
  - WRF- As run by NCEP (NOAA)
    - 12km resolution at 60 vertical levels
  - MM5- As run by AFWA
    - Nested Grid 45km, 15km resolution at 42 vertical levels
    - 5km output over selected regions



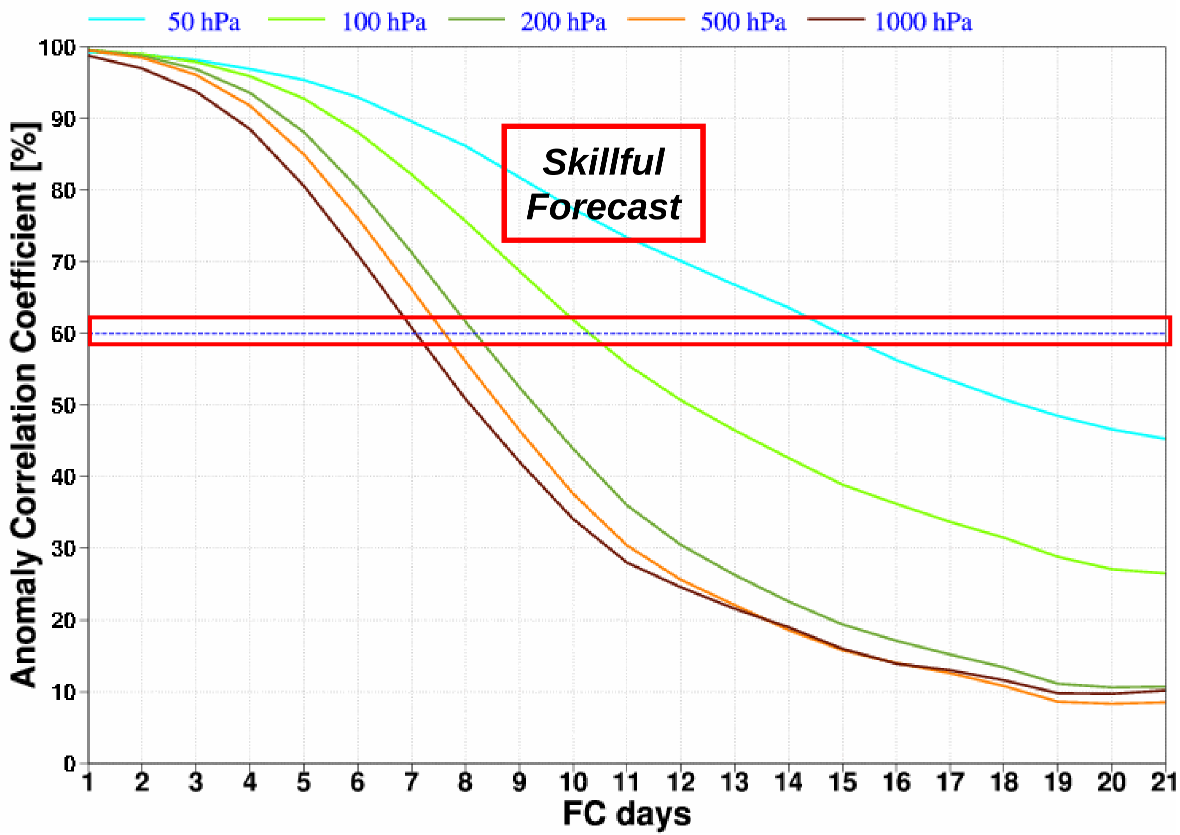
https://weather.afwa.af.mil/data_links/MWORLDLOCA.GIF

NWP Forecast Skill Improvement



Anomaly Correlation of 500-hPa height for 3-, 5-, and 7-day forecasts for the ECMWF operational model as a function of year. Top and bottom of each band correspond to Northern and Southern Hemispheres, respectively.

- **Difference between hemispheres has nearly disappeared in the past few years due to the successful assimilation of satellite data**



Anomaly Correlation Coefficient for ECMWF forecasts for different levels over the Northern Hemisphere in 2004.

http://www.ecmwf.int/products/forecasts/guide/The_data_assimilation_and_analysis_system.html

Introduction to Numerical Weather Prediction

- **Sensitivity to parameterization of sub-grid processes**
  - **Introduction of chaos and error into models**

- **Model has O(10^7) values to calculate (degrees of freedom) and observational quantities are O(10^5)**

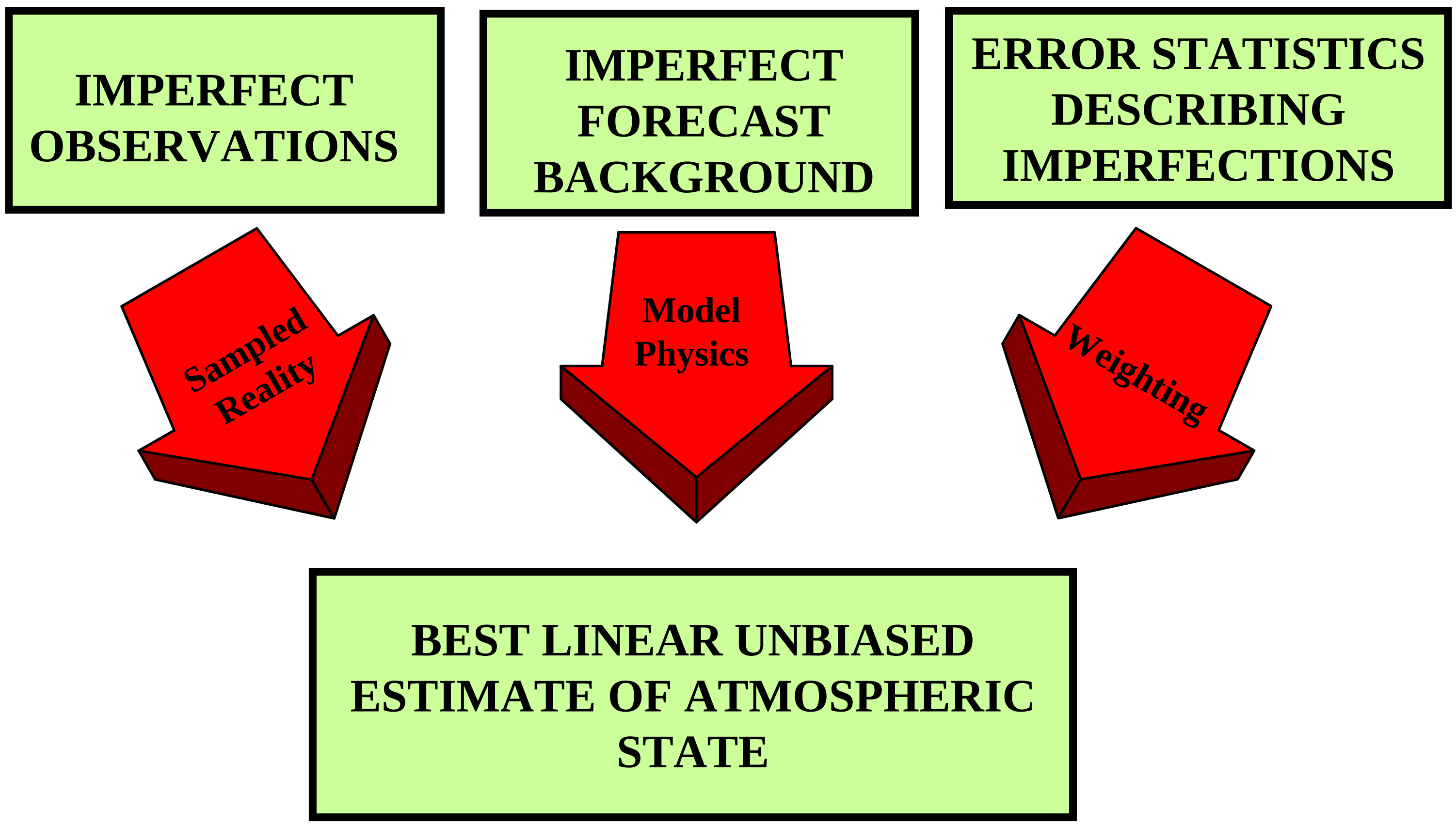
- **Problem statement:** ***NWP is an under-determined initial value problem; how do we obtain the most realistic representation of the initial condition?***

Three Atmospheres

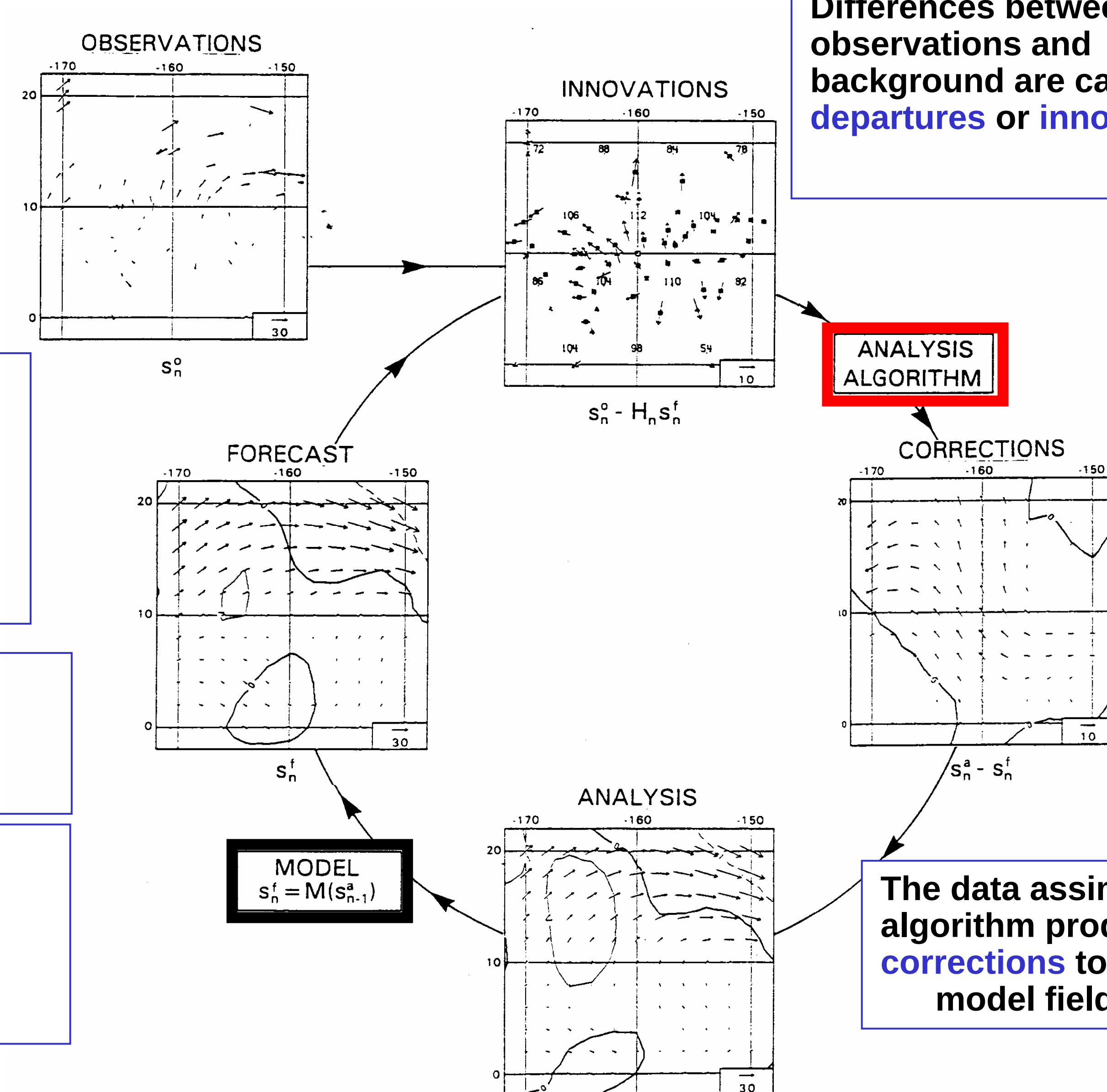
We deal with three atmospheres

- **Real atmosphere**
  - Unknown
- **Observed atmosphere**
  - Data coverage gaps
    —vertical, horizontal, temporal
  - Observation error
    —random and systematic
- **Analysis/model atmosphere**
  - Observation limitations
  - Data assimilation system limitations
  - Model limitations

Three Sources of Information



Analysis/Forecast Cycle



Differences between observations and background are called departures or innovations

The forecast model provides the background estimate of the current atmospheric state

The analysis provides the initial conditions for the next forecast

These corrections are added to the background to form the analysis

The data assimilation algorithm produces corrections to the model fields

Methods of Data Assimilation

- Optimal Interpolation (OI)
- 3-Dimensional Variational Assimilation (3DVAR)
- 4-Dimensional Variational Assimilation (4DVAR)
- Extended Kalman Filter
- Ensemble Kalman Filter

Methods of Data Assimilation

- **Optimal Interpolation**

one observation

one background value

$$T_a = T_b + W(T_o - T_b)$$

multiple variables $\Rightarrow$

interpolate each grid point to the observation location first

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_0 - \mathrm{H}(\mathbf{x}_b)]$$

then weight each innovation

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}\,\mathbf{d}$$

number of model variables

number of observations

$$\mathbf{x}_a = \begin{bmatrix} x_1 \\ \vdots \\ n \\ \vdots \\ x_{10^7} \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} d_1 \\ \vdots \\ p \\ \vdots \\ d_{10^6} \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w_1 & \cdots & w_{p1} \\ \vdots & & \vdots \\ \vdots & n \times p & \vdots \\ \vdots & & \vdots \\ w_{1n} & \cdots & w_{np} \end{bmatrix}$$

Methods of Data Assimilation

- **Optimal Interpolation**
  - **Optimal analysis is found by minimizing the analysis error variance (A) by finding the optimal weights of observation increments through a least squares approach**

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$$

$$W = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}$$

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$

Optimized

$$\mathbf{A} = (\mathbf{I} - \mathbf{W}\mathbf{H})\,\mathbf{B}$$

Minimized

σ_a^2 = analysis error variance
σ_o^2 = observational error variance
σ_b^2 = background error variance
W = optimal weight
W = weight matrix
B = background error covariance matrix
R = observations error covariance matrix
H = linear observation operator matrix
I = identity matrix
A = analysis error covariance matrix

Methods of Data Assimilation

- **3DVAR Introduction- Cost Function**

Why use the term Cost Function?

$$J(x) = \frac{1}{2}\left[\frac{(y-x_a)^2}{\sigma_o^{\ 2}} + \frac{(x_b-x_a)^2}{\sigma_b^{\ 2}}\right]$$

x_a = analyzed value
y = observed value
x_b = background value

Think of the function as defining how far away the analysis value is away from the input values. You will be "charged" for not fitting the observation. You will also be "charged" for not fitting the background. You choose the analysis to minimize your "cost."

Methods of Data Assimilation

- **3DVAR Introduction- Vector Form of Cost Function**

Vector Form:

$$J(x) \quad = \quad \frac{1}{2}\left[(y-x_a)\mathbf{R}^{-1}(y-x_a)+(x_b-x_a)\mathbf{B}^{-1}(x_b-x_a)]\right]$$

Minimizing *J(x)* with respect to x_a yields:

$$x_a = x_b + \mathbf{B}(\mathbf{B}+\mathbf{R})^{-1}(y-x_b)$$

Methods of Data Assimilation

- **3DVAR**
  - **Minimizes cost function**
  - **Performs analysis at set times**
    - **Only observational data from set times are included or observations in a +/- time window are lumped together and given essentially an equivalent time**
  - **Employs a forward operator or observation operator: H**
    - **Interpolates observations spatially to grid points**
    - **Converts observed quantities (i.e. satellite radiances, radar reflectivities) into model variables (i.e. temperatures, humidites)**

$$J(x) \quad = \quad \min \frac{1}{2}\left[\underbrace{(x_b - x_a)\mathbf{B}^{-1}(x_b - x_a)}_{\text{Distance to forecast}} + \underbrace{(y - \mathbf{H}x_a)\mathbf{R}^{-1}(y - \mathbf{H}x_a)}_{\text{Distance to observations}}\right]$$

At analysis time

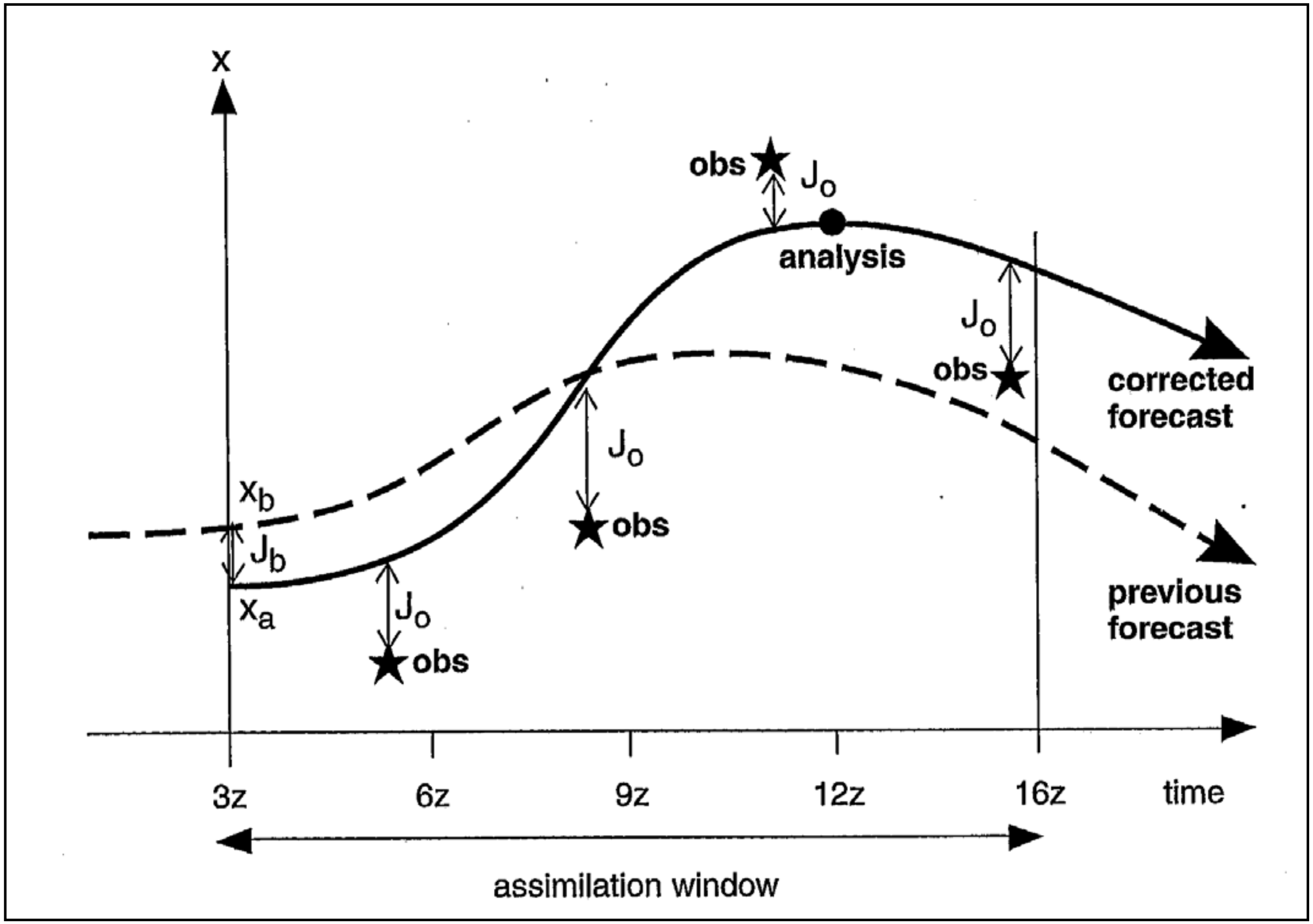
Methods of Data Assimilation

- 4DVAR
  - Minimizes cost function
  - Cost function is weighted difference between model forecasts during an assimilation window and coincident observations

$$J(x) \quad = \quad \min \frac{1}{2}\left[\underbrace{(x(t_0)-x_a(t_0))\mathbf{B}^{-1}(x_b(t_0)-x_a(t_0))}_{\text{Distance to background at initial time}} + \sum_{i=0}^{N} \underbrace{(y_i-\mathbf{H}x_i)\mathbf{R}^{-1}(y_i-\mathbf{H}x_a)}_{\text{Distance to observations in a time window interval}} \right]$$

  - 4D-Var seeks initial conditions such that the forecast best fits the observations within the assimilation interval
  - Computes the increments at observation times during a forward integration using the forecast model and then integrates these weighted increments back into the initial time using the adjoint model ($\mathbf{L}^T$)

Methods of Data Assimilation



4DVAR data assimilation window

Methods of Data Assimilation

- **Extended Kalman Filter**
  - **Same weighted approach as OI, but using a background error covariance matrix (P) that evolves with the forecast rather than a constant background covariance matrix (B)**
  - **Forecast error covariance is obtained using a tangent linear model (L) to transform the perturbation from the initial time to the final time and the adjoint model (L^T) to "advance" the perturbation backwards from the final to initial time to optimize the initial conditions**

Forecast step:

$$x_n^f = M_n(x_{n-1}^a)$$

$$\mathbf{P}_n = \mathbf{L}_n \mathbf{A}_{n-1} \mathbf{L}_n^T + \mathbf{Q}_n$$

Analysis step:

$$x_n^a = x_n^f + \mathbf{K}_n (y_n - H x_n^f)$$

The optimal weight matrix :

$$\mathbf{K}_n = \mathbf{P}_n (\mathbf{R} + \mathbf{H}\mathbf{P}_n \mathbf{H}^T)^{-1}$$

The analysis error covariance:

$$\mathbf{A}_n = (\mathbf{I} - \mathbf{K}_n \mathbf{H})_n$$

Methods of Data Assimilation

- “Extended Kalman filter is gold standard of data assimilation” (Kalnay, 2006)

- A poor initial guess can be transitioned through time to provide the best linear unbiased estimate of the state of the atmosphere and its error covariance provided observations are frequent and system is stable (Kalnay, 2006)

- Very computationally expensive

- So can it be done more cheaply?

Methods of Data Assimilation

- **Ensemble Kalman Filter**
  - **Estimates forecast error covariance matrix from ensemble of forecasts initialized by random perturbations added to the same sets of observations**
  - **Computational cost increased $O(10^2)$ from OI or 3DVAR, but cheap compared to Extended Kalman Filter whose cost is $O(10^7)$**
  - **Tangent linear or adjoint model not necessary**
  - **Most promising approach for the future**

Conclusions

- Limit of predictability is ~2 weeks due to sensitive dependence on initial condition (chaos)
- Quality of forecast therefore is critically dependent on the quality of the initial condition
- Throughout 50 years of NWP, methods of obtaining that initial condition have evolved with increasing sophistication of NWP models, new remotely-sensed observation types, and increasing computing power

Acknowledgments

- Holton, James R., 2004: *An Introduction to Dynamic Meteorology Fourth Edition.* Elsevier Academic Press, Burlington, MA.
- Kalnay, Eugenia, 2006: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, New York.
- Weygandt, Stephen S., 2006: *Assessing The Impact Of Current And Future Observing Systems in Environmental Predictions.* NOAA Earth System Research Laboratory

- Slides 9-15 and 21-25 were taken directly from course notes and presentations of MR4323, Air Ocean Numerical Modeling, at the Naval Postgraduate School. A special thanks to Commander Rebecca Stone and Prof Mary Jordan for providing this material.

Questions?